

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/117063/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Jeanes, Annette, Coen, Pietro, Drey, Nicholas and Gould, Dinah 2019. The validity of hand hygiene compliance measurement by observation: a critical systematic review. *American Journal of Infection Control* 47 (1) , pp. 313-322.

Publishers page: <https://doi.org/10.1016/j.ajic.2018.08.004>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# **The validity of hand hygiene compliance measurement by observation: a critical systematic review**

**A Jeanes<sup>1</sup>, PG Coen<sup>1</sup>, NS Drey<sup>2</sup> & DJ Gould<sup>3</sup>**

**<sup>1</sup>Infection Control Department  
University College London Hospitals  
3<sup>rd</sup> floor East  
250 Euston Road  
London NW1 2PG**

**Tel: 0203 447 8974**

**Email: [annette.jeanes5@gmail.com](mailto:annette.jeanes5@gmail.com)  
[Annette.jeanes@uclh.nhs.uk](mailto:Annette.jeanes@uclh.nhs.uk)**

**<sup>2</sup>City University London  
Northampton Square  
London EC1V OHB**

**<sup>3</sup>Cardiff University  
4th Floor,  
Eastgate House,  
35-43 Newport Road,  
Cardiff CF24 0AB**

**Word count :**

**Abstract 248**

**Main document :4814**

**1 diagram, 1 graph, 7 tables**

## **Summary**

### ***Background***

To improve and sustain compliance, hand hygiene practice is monitored by observing practice. There are threats to the validity of the data collected by direct observation.

### ***Aims***

To identify and describe the potential biases in hand hygiene compliance monitoring by direct observation

Describe the typology of these biases

Propose improvements to reduce bias and increase the validity of this methodology

### ***Methods***

A systematic review of hand hygiene intervention studies assessed for the presence and type of methodological bias.

### ***Findings***

There was inconsistency of terminology, definitions, criteria, tools and description of the data collection methodology. Frequency of observation and/or study length was not stated or unclear in 54 (76%) publications. The observers were trained in 55 (77%) publications although this varied in content and was only clearly specified in 23 (32%) publications.

None of the 71 studies reviewed were free of bias. Selection bias included lack of week-end measurement in 61 (86%) studies and lack of night time measurement in 46 (65%) studies. Observations were undertaken in single-specialty ward locations such as adult and neonatal intensive care or paediatrics in 35 (49%) studies.

Inter-rater reliability was undertaken in 26 (37%) studies but only 6 (8%) undertook this as an on- going quality assurance process rather than just in training.

### ***Conclusion***

Measuring hand hygiene compliance by direct observation lacks validity. Regular monitoring of barriers to compliance could identify areas for improvement to improve compliance. To enable comparison and evaluation of studies, the methodology should be reported in a standardised format. (248)

***Key words:*** Hand hygiene, observation, validity, Hawthorne effect,

## Introduction

Historically hand hygiene compliance (HHC) in healthcare has been poor<sup>1,2</sup>, despite evidence that it can reduce the risk of infection<sup>3</sup>. To improve and sustain compliance, regular hand HHC monitoring in health care organisations is recommended<sup>4</sup>. This data is also used to provide assurance of compliance and performance.

Direct observation of HHC is widely accepted as the ‘gold standard’<sup>5,6</sup> but there are threats to validity associated with the collection of data by observing humans<sup>7</sup>. This includes the ‘Hawthorne effect’<sup>8</sup> which is a reactive response to being observed or studied<sup>9</sup>. This may increase hand hygiene frequency during observation<sup>10</sup>.

HHC data quality has been criticised<sup>11</sup> and the current ‘gold standard’ is regarded by many as inaccurate<sup>12,13,14,15</sup>. The provision of robust and credible data to measure performance is important in promoting and sustaining evidence based practice and quality improvements<sup>16</sup>.

Gould et al<sup>17</sup> found insufficient information about the data collection methods to assess the reliability of data in a review of 42 hand hygiene interventional studies. They found limited consideration of the Hawthorne effect, observer training, inter-rater reliability and corroborative data collection. In addition, there was sampling bias including day time observation in intensive care units.

Haas and Larson<sup>18</sup> reviewed 31 hand hygiene adherence studies. They found limited details of observer training, variations in the definitions of compliance, small sample size, sample selection bias including a lack of night time sampling and the presence of the Hawthorn effect. They commented that observers monitoring HH compliance within their workplace, were likely to be biased and recognised. These findings were echoed in a review of compliance monitoring by Boyce<sup>19</sup> who also found that data collection tools varied.

Harrington et al<sup>20</sup> examined 28 studies and 2 US hand hygiene guidelines from 1990-2006 to find a measure of hand hygiene which was reliable and valid. Only 19 (68%) of the publications addressed reliability and validity and there was variation in methods, tools and the approach to validity where it was reported.

In a review of 96 studies Erasmus et al<sup>21</sup> found that the definitions of hand hygiene expectations and compliance varied between studies. They concluded that these issues could be resolved in part by a standardised ‘*measuring instrument*’ and standardised reporting to facilitate the comparison of data.

The risk of bias was part of the rationale for the development of the WHO hand hygiene observation method and tool<sup>22,23</sup>. However, this was not universally utilised. Ellingson et al<sup>24</sup> identified continued variation in practice which included monitoring tools and definitions of compliance. Ellingson et al commented that there was now

further complexity as observation with rapid feedback was becoming part of the improvement intervention.

Research methods and design affect the validity of research findings<sup>25</sup>. In simple terms validity is about ensuring we are measuring what we think we are measuring<sup>26</sup>. There are two categories of validity: internal and external validity<sup>27</sup>. External validity relates to the generalizability or extrapolation of results<sup>28</sup>. Internal validity is the ability to accurately measure what is required whilst avoiding bias or error. The key types of bias associated with HHC compliance measurement are information<sup>29</sup>, selection<sup>30,31</sup> and confounding bias<sup>32</sup>.

### **Information including measurement bias**

The Hawthorne effect<sup>8</sup> has been identified in several HHC studies<sup>33, 34,35,36,37</sup>. The 'novelty effect' of being observed may diminish with time<sup>38</sup> particularly if it is routine<sup>39</sup>. However, the presence of people who are known HH monitors may act as a prompt for improved HHC<sup>40</sup>. Overt compared to covert observation has been associated with an increase in measured HHC<sup>33, 36, 37</sup>. Estimates of the effect vary but it could inflate HHC performance by between 30-50%<sup>36, 37</sup>.

Allegiance and peer pressure may be associated with bias of the observer<sup>36,40</sup>. In addition, forms of confirmation bias<sup>41</sup> where interpretation of information is affected by pre-existing beliefs may influence results particularly when researchers are observers<sup>42</sup> or when observers are not blinded to the hypothesis<sup>37</sup>. Unconscious bias may prompt selective observation where attention is focused on data supporting the hypothesis<sup>43</sup> for example ignoring cleaners HHC based on a belief that their role in transmission is insignificant.

Employing several observers may reduce the effect of individual bias but increases inter observer variability<sup>44</sup>. Undertaking regular inter-rater scoring may assist in controlling this issue.

The scoring precision and consistency of the observer may be affected by numerous factors including perception, training, experience, fatigue, length of study<sup>45</sup> and the data collection tool. Accuracy may improve with training and practice<sup>46</sup>. When combined with an optimal data collection instrument, clear instructions and definitions this may also reduce observer drift<sup>45</sup> and inter-rater variability.

Recording rapid successive actions may lead to errors in recording as it is 'impossible to capture everything'<sup>47</sup>. In addition, errors may be related to observer fatigue which may occur when undertaking prolonged periods of observation<sup>48</sup>.

Public reporting of HHC data may affect the data produced particularly when linked with performance rewards<sup>49</sup>. Funding or sponsorship for a research study may also prompt those with a vested interest to seek positive results.

### **Selection bias**

Clinical setting may introduce bias. In HHC monitoring many studies focus on intensive or neonatal care units which are not representative of all healthcare settings<sup>1, 2, 50-57</sup>. Their configuration may offer greater visibility and ease of observation compared to more enclosed environments where visibility is limited<sup>36</sup>. To reduce bias, it is preferable to sample a wider selection of the population studied to ensure it is more representative<sup>58</sup>. Self-selection may introduce bias<sup>59</sup> when consent to participate is required as those being observed can opt in or out of studies.

There is also potential for bias when some observers undertake more audits than other observers; when observers differ in their ability to measure; this can be dealt with by multi-level regression analysis, which is invariably done in combination with multivariate analysis. This relies on the ability to record observer identifiers in the audit records.

The ability of the observer to witness clinical practice is essential and therefore generally excludes accurate observation at night. Whilst observing what is happening behind curtains, in single rooms or consulting rooms requires the presence of an observer which may affect patient privacy and dignity.

Weekends and nights may have less staff and activity than week days. The selection and use of differing time periods, staffing and activity may result in systematic error in data collection<sup>58</sup> and the results will be unrepresentative. In addition, the capacity of staff to observe may be confined to quiet periods which may not reflect behaviours during busier periods<sup>60</sup>.

Ad hoc samples may also be unrepresentative compared to regular planned sampling. Continuous sampling may be more reliable than intermittent sampling as the latter is more likely to contain omission errors<sup>61</sup>.

### **Confounding bias**

Confounding bias can create an outcome which in turn influences the results<sup>29</sup>. For example, the finding that HHC in doctors is lower than nurses may be biased as (until recently) doctors were more likely to be male, a group which have lower hand hygiene compliance than females. Male gender confounds the result.

However confounding bias relates to the ability of the data collector to collate potential confounder information which can be used in a multivariate analysis to minimise bias. Multi-level analysis can, in addition, reduce estimation bias when there are differences in the audit workload across observers.

It is important to control for confounding bias when studies are designed and remain alert to its possible existence during analysis and interpretation of findings. In the example above, confounding bias could be avoided by matching staff in terms of occupational group and gender. Failure to do so would result in lack of comparability between audits.

The validity of this information is important as these data influence organisational risk perception, resource allocation and ultimately patient outcomes. HHC measurement takes time which uses valuable resource but can indicate areas requiring improvement. There is a risk of selection, information and confounding bias in the direct observation of HHC which threatens the internal and external validity of the data collected <sup>26</sup>.

To improve practice, clinicians require valid data from a credible source<sup>62</sup>. As the strength of the evidence provided, influences the success of change implementation<sup>63</sup>, understanding issues of validity related to the use of the current method will assist in developing improvements or changes in this process.

Though hand hygiene compliance studies have been reviewed previously, the validity and the potential for bias of studies has not been reviewed in detail. The validity of HHC by observation is the focus of this review.

### **Research questions of this systematic review**

1. Is there evidence that HHC monitoring by direct observation lacks validity?
2. If there is evidence of bias:
  - a. In what ways are studies which have used observation to measure HHC biased?
  - b. What is the taxonomy of bias in publications
3. How could the validity of HHC measurement by direct observation be improved?

## **Methods**

### **Inclusion and exclusion criteria**

The search strategy identified studies employing direct observation to monitor HHC of health workers.

Published peer reviewed full text studies and reports worldwide were eligible. Papers with no published abstract were excluded as it was not possible to assess these papers against the inclusion/exclusion criteria or extract the data required.

Publications prior to 1970 were excluded on the basis that most hand hygiene monitoring was associated with improving compliance and this was established after this date. Publications up to 2015 were included.

Only publications which reported measured observed hand hygiene compliance in health care facilities for humans were included. Systems such as video monitoring and electronic monitoring were excluded as they are not comparable with direct human observation of humans. Studies with multiple interventions were included if HHC by direct observation was included as a component.

Surveys of beliefs/self-reporting, modelling, reviews, opinion pieces, poster submissions, commentaries including systematic reviews, rural studies, animals, vets, zoos, catering, experiments or simulations of HHC, indirect monitoring such as soap usage, visitor compliance or visitors/patients measuring HCW compliance were excluded.

### **Information sources**

The literature was searched using BNI, PubMed, Scopus, MEDLINE, Health Business Elite and CINAHL databases via NHS Athens. In addition, the work of key authors in the field was identified, grey literature primarily from NHS portals was reviewed, suggestions from other experts were sought, a hand search of current relevant literature was undertaken and an alert of news items including published papers relating to hand hygiene compliance via Google was activated.

Initially the systematic reviews of Haas & Larson 2007<sup>64</sup>, Gould et al 2008<sup>65</sup>, Gould et al 2010<sup>66</sup>, Erasmus et al 2010<sup>21</sup>, Huis et al 2012<sup>67</sup>, were examined and key terms used from these publications informed terms used in the search strategy.

Key terms used in the search included: 'hand hygiene', 'hand hygiene compliance', 'staff' 'observation', 'assurance', 'compliance monitoring', 'compliance measurement', 'performance monitoring', 'performance measurement', 'quality improvement' 'audit', 'reporting', 'interpreting/interpretation', 'direct observation', 'feedback', 'competence', 'knowledge', '5 (five) moments', 'behaviour', 'reliability' 'validity' 'accuracy' and 'hand wash/washing', 'clean hands'. Terms were used in combination.

Subsequently results were checked to ensure the key authors literature had been identified in the search.

Extract from a search undertaken 1st September 2015 summarised in Table I

Limits applied: Full Text; Published Date: 01/ 01/ 1989-31/12/ 2014; English Language, Search modes - Boolean/Phrase via Interface of EBSCOhost Research Databases Health Business Elite; CINAHL with Full Text; MEDLINE.

### **Study selection**



Each paper was initially examined to ensure the description of HHC monitoring included sufficient detail to extract data. We looked for details that allow assessment of methodological bias. A formal assessment of repeatability<sup>68</sup> was not possible as the data in publications was not comparable or absent.

Reviewer (AJ) undertook the initial search and applied the exclusion criteria to abstracts. This was repeated with the remaining full text publications. Then with the second reviewer (PC) the remaining selected studies were reviewed to ensure they met the inclusion criteria.

The study size and outcome of the intervention or measurement were not important in examining the methodology and were not factors in the data collection or selection of the publications.

The study characteristics required for inclusion were:

- Measurement of HHC of HCWs by direct observation, in acute healthcare settings.
- The reported methodology provided sufficient information to determine the validity of methods. This was defined as the presence of sufficient information within the publication to complete the data set.
- The characteristics from which data were extracted is summarised in the Data collection of bias and rational for inclusion table (Table II)

More than 5,000 publications were initially identified. The application of exclusion criteria yielded 118 publications. 47 full text publications were then excluded following independent review. (Diagram I)

The most frequent reason for exclusion in 31 publications (66%) was unclear detail of the observation method. This is summarised in table III

### **Data extraction**

An EXCEL spread sheet data collection tool was the agreed and tested. The information sought in the data extraction tool reflected current literature from a range of scientific research relating to bias. (Table II)

Each reviewer then independently extracted data from each study. This included the purpose of the study, length of the study, methods and tool used, definitions utilized, reliability and validity of the approach and clarity of the description. The quality of the clarity of the methodology reported was assessed in terms of how easy it would be to replicate the study using the information provided. The data collected were then compared and any mismatches were discussed and resolved by agreement.

Some of the detail about the methods used in the data collection was missing or unclear e.g. frequency of observation. It was agreed that these papers would not be excluded if they contained a minimum data set. The rational was to provide a representative sample of studies.

Meta-analysis of the studies was not appropriate in this instance as the focus of the review was upon the methods employed rather than outcomes.

### **Assessment of bias**

An assessment of bias of the publications selected was undertaken to indicate a measure of the extent of bias present. A bias scoring system was developed in the absence of available and appropriate tools in the literature. The assessment was based on 15 potential bias components associated with HHC from literature. This included 6 selection, 7 information and 2 confounding bias components. Each of the criteria was equally weighted and the maximum bias score was 15. The bias sub-scores are in Table IV.

The following assumptions formed the basis of the assessment (bias was assumed whenever the details provided in the paper did not answer the question):

1. **Who was observed?**The observation of one staff group produces selection bias
2. **Which specialities were observed?**The observation of a single speciality produces selection bias
3. **How many hospitals were included?**The assessment of >1 hospital reduces selection bias
4. **Did they also monitor nights?**The inclusion of night time observation reduces selection bias
5. **Did they monitor week-ends?** The inclusion of week-end observation reduces selection bias
6. **Was informed consent required?**Providing informed consent may increase awareness of observation and may increase information bias
7. **Was author of paper an observer?**The author may have a vested interest which increases information bias
8. **How many auditors?**One auditor may increase information bias whilst more will reduce bias
9. **Were auditors trained?**Training may reduce information bias
10. **Were observations undertaken in a consistent frequency?**Inconsistent periods of time may increase information bias
11. **Was inter-observer variation measured?**Presence of tests of inter- observer variation indicates awareness of potential for information bias
12. **Were observers internal or external?**Internal observers may have allegiance or be susceptible to other pressures and increase the risk of information bias
13. **Was a multivariate analysis undertaken?** A multivariate analysis reduces the risk of confounding bias (assuming all relevant potential confounders have been measured and included in the multivariate model)
14. **Was a cluster analysis undertaken?** A cluster analysis reduces the risk of selection bias because it deals with the problem of unequal distribution of HHOs among observers.
15. **Was it a long or short study?** A long study (>1 year) increases the risk of a deterioration in accuracy and changes in staff which may contribute to information bias increase confounding bias; it also minimises the impact of seasonal bias (all seasons being covered) which reduces confounding bias.

Other elements which may affect bias were not included in the bias assessment. The reasons are briefly described.

The reason for measurement was unclear in some studies and in some others, it was inconsistent within the description given. An attempt to definitively categorise the studies as intervention, baseline or tool development data collection studies was unsuccessful.

Pilot studies were undertaken in 16 (22%) studies<sup>72,73,78,80,86,104,90,113,114,122,93,131,132,123,92,94</sup> but it was unclear if these were undertaken to identify and reduce bias or to assess feasibility.

Covert, overt, obtrusive, unobtrusive observation were excluded as the definitions appeared to vary. This included managers and leaders aware of covert observation<sup>78,87,88</sup>, covert and overt in one study<sup>34</sup>, covert with feedback<sup>82</sup> or it was unclear.

## **Data analysis**

We used STATA 12 for statistical analysis of the data extracted. For significance tests we used Fisher's exact test. We used Pearson's correlation coefficient to investigate the correlation matrix including the Bonferroni adjustment of p-values in inferring their significance. We tested for deviation from normality for the summary bias score by means of the skewness/kurtosis test, the Shapiro-Wilk W test and the Box-Cox regression model.

## **Results**

### **Study selection**

71 publications were included in the review<sup>34, 69-137</sup>

### **Publication selection bias**

No significant trends were detected in the proportion included by country of origin ( $p = 0.259$ ) and the year of publication ( $p = 0.188$ ; Fisher's exact test).

The selected and excluded publications were from a wide range of countries. Most were from Europe and North America.

### **Presence of bias in included publications**

There was evidence of potential bias in all the 71 included papers, though the degree of bias was variable. Table V is a summary of bias in studies. The bias components are summarised in table VI, the frequency in Table VII.

## **Bias within the literature**

## Information bias including measurement bias

The Hawthorne effect<sup>8</sup> was identified by authors in 12 (17%) studies. 31 (44%) studies attempted to control for this bias by covert or inconspicuous observation; an example was partial concealment of observer by placement in the corner of the area<sup>80</sup>. One study was halted when staff became suspicious of observers<sup>106</sup>.

In some studies, the role of the observers may have become clear to the staff observed as observers followed the HCW<sup>94, 98, 100</sup>, provided individual feedback<sup>127</sup> or were in the patient's room<sup>83, 85</sup>. There were also prolonged observation periods<sup>113, 114</sup>, observation of one person for 2 hours per shift on three occasions<sup>76</sup> and two observers who observed the same person simultaneously<sup>77</sup>. Though only 11 (15%) of studies required informed consent from staff, 41 (58%) publications required ethics or a similar approval process which potentially increases awareness of observation. In one study, it was noted that compliance increased the longer auditors remained in the area<sup>133</sup>.

The number of observers was not stated in 31 (44%) publications but were most frequently 1-2 people. Observers were trained in 55 (77%) studies. The training varied and included written instructions, DVD/video, lectures, workshops, scenarios, simulations, familiarisation and concurrent pilot or trial observations. In 9 (13%) studies, observers had received training previously. The method of training was only specified clearly in 23 (32%) of studies. Validation of scoring within training was undertaken in 28 (39%) of studies

15 (21%) studies reported observers were internal, in 11 (16%) they were external to the organisation, but in 45 (63%) it was not stated or unclear if observers were external or internal. In 12 studies (17%) authors/researchers were observers.

In 47 (66%) studies the duration of the study was <twelve months and 18 (25%) were >twelve months. Length of the observation period varied. In 18 (25%) studies observation was <one hour, in 16 (23%) it was >one hour including some continuous measurement for 24 hours<sup>113, 114</sup>.

Audit frequency and/or study length were not stated or unclear in 54 (76%) of studies. The frequency of observation measurement was clearly stated in 16 (23%) studies.

Inter-rater reliability was undertaken in 26 (37%) studies though for 16 studies this only took place in training. There were on-going tests such as kappa, for inter-rater reliability in only 6 (8%) of studies.

Assessment of information bias was hampered by lack of details of methods. However, validity may have been improved by training auditors, regular auditing for consistent periods of time and validation of scores beyond the training period.

## Selection Bias

Sampling bias was evident in all studies.

Observations were undertaken in single-specialty ward locations such as adult and neonatal intensive care or paediatrics in 35 (49%) studies. Most monitoring was undertaken in one hospital with 11 (16%) monitoring in more than one hospital.

50 (70%) of studies reported observation was done partly or entirely in the day time. Observation at night was undertaken in 25 (35%) studies, whilst weekend observation was undertaken in 10 (14%) studies.

The health care workers observed were primarily doctors and nurses though in 16 (23%) studies the role of the HCW was not specified. The people undertaking the observation was unspecified in 33 (46%) papers but otherwise included students, infection control staff, nurses, researchers and doctors.

## **Confounding Bias**

44 (62%) studies did not undertake a multivariate analysis and 55 (77%) were short studies i.e. < one year or duration not stated.

### **Measurements in addition to observation**

20 (28%) studies measured hand hygiene products though the method varied and included staff assessing how much was left in individual dispensers<sup>78,130</sup>. 16 (23%) assessed hand hygiene method which variously included time taken, coverage of hands, drying, turning off taps.

### **Overall rating of bias**

The result of the summary bias score was available for all the 71 selected studies with a median of 8 units (2 to 13 range), a mean of 8.25 (2.05 standard deviation) – no study scored zero or one. The distribution of scores did not deviate significantly from normality and none of the sub-scores were significantly correlated to any other, suggesting that no sub-score was a redundant correlate of any other. (Graph I)

## **Comparability of publications study findings**

Data from the different studies was not comparable as definitions, terminology, measurement criteria and methodologies varied.

Most studies did not specify how they undertook the observations in detail. The number of observations undertaken or other outcome measures was not reported or was unclear in several studies<sup>81,95,124,125,132,133,136,137</sup> and in others the data was not comparable. The periods of time observed, the number of areas observed during the observation varied considerably and were not comparable across studies.

Definitions of HH and HHO were unclear in some studies<sup>73,78,79,96</sup>, in others definitions were variably applied. In describing hand hygiene measurement at least 60 terms were used. Alcohol hand decontaminants alone accounted for seven terms.

Several studies reported that they used standard tools such as the WHO compliance tool but many created their own<sup>70,73,75,78,81,86,92,97,90,2,91,124,130,135</sup>. Others modified or adapted the tools used<sup>74,79,80,119,34,83,87,100,104,105,118,125</sup>. The WHO guidelines were modified to the Six moments<sup>98</sup> and Four moments of hand hygiene<sup>76,77,137</sup>.

The variation and adjustment in tools made summarising and comparing the criteria used for measurement difficult. Examples included Boscart et al<sup>76</sup> which used the 'Ontario tool' in which the WHO moments of "after-patient-contact" and "after contact with patient environment" are combined and "before patient contact" and "before contact with patient environment" are also combined.

The hand hygiene expectation associated with glove use was ambiguous. In some tools failure to perform hand hygiene after glove removal was considered non-compliant<sup>78,79,90,132</sup> whilst in others glove use it was not included in monitoring.

Other differences included only stipulating hand hygiene following contact with a contaminated environment or objects, rather than all patient environments<sup>80, 97</sup>. This extended to applying a risk assessment to criteria in some studies<sup>107,109</sup>.

Some used very specific actions and expectations<sup>103,116,129</sup> whilst others referred to standard criteria such as 'WHO 5 moments'. Others were specific but omitted to explain if the expectation was before and after the contact<sup>124</sup>. Other adjustments included, excluding the first patient contact because observers were waiting outside the patient room and could not see if the HCW cleaned their hands in the previous room<sup>108</sup>, whilst others focused only on hand hygiene before contact with the patient as it was perceived to be important and to simplify the observers' task<sup>118</sup>.

Inconsistencies between studies made direct comparisons difficult. However, the studies were written without a standardised expectation of output.

## **Limitations of the review**

Great emphasis has been placed on HHC in recent years, resulting in the publication of large numbers of studies in a wide range of journals. It is therefore possible that some might not have been detected during the searches.

Despite the exclusion of many publications based on the lack of sufficient detail in the methodology to extract the detailed data required, it became clear that many of the studies selected were not designed for this specific data extraction.

A major limitation was the inability to identify or measure the potential Hawthorne effect in studies. Though this has been recognised and described previously in HHC<sup>36, 37</sup>.

## **Conclusions**

This review includes a comprehensive consideration of bias which is more detailed than previous reviews. None of the studies reviewed could be considered bias free and

the methods employed could therefore not be considered to provide a valid assessment of HHC. The types of bias identified reflects those found in previous reviews<sup>17,18,19,20,21</sup>.

Few of the selected studies provided a clear description of the data collection methodology. This is surprising as replication allows confirmation of findings<sup>68</sup> and comparison. An expectation of a peer review is that the methodology is scrutinised to determine the quality of the study and results.

Standardisation reporting these studies has been recommended previously<sup>17,19</sup>. This would enable replication and comparison with other studies. This should include details of how the observation was undertaken, length of observation, frequency, training of observers, inter-observer validation, definitions, tool used and any adaptations of tools.

Standardisation of training and inter-rater measurement would also be beneficial. The assumption that people know how to observe, may be unfounded. It enables repetition of previous methodological and observer bias. All observers should receive training prior to commencing measurement and regular external validation to support consistent standards and practice.

The Hawthorne effect was acknowledged in several studies. Though a constant threat to the validity of the data collected, the Hawthorne effect could be viewed as a systematic error in the observational methodology which is relatively constant and error tolerance could be applied. The data collected is a sample of behaviour which will be affected by several variables. Though potentially inaccurate, if the methods, conditions and degree of error are relatively constant, then the results of observation may be a pragmatic indicator of performance for inspection of trends.

There are benefits in observing practice, including improving practice<sup>138</sup>. Observation is used to assess clinical competence<sup>139</sup> and to gain insights into what happens in practice. Klien<sup>140</sup> suggests that developing insight may require rejection or modification of established assumptions to develop a new approach to issues. 'Gold standards' may be challenged and other potential solutions or ideas may be generated.

A structured and systematic approach to observation would be more rigorous and reproducible than random observations. However, limiting or ignoring what could be useful to follow up and restricting observations to a predetermined and rigid format, may miss important serendipitous findings. Repeatedly just observing HHC may inadvertently create blindness to other significant events as attention may be highly selective<sup>141</sup>. Even experienced observers may be subject to in-attentional blindness when focused on a single process which is familiar and predictable<sup>142</sup>.

However, experienced observers may be more successful than a novice at detecting patterns and anomalies<sup>143</sup>. Expertise and preparedness create a 'search image' which combined with situational awareness filters out irrelevant information which may overwhelm the analytical skills of a novice<sup>144</sup>. Observation by someone with relevant experience and training and education could be beneficial.

Though some studies checked product usage and hand hygiene method, apart from education, potential barriers to compliance were rarely noted. The identification of barriers and utilisation of improvement opportunities, could add value to the monitoring process. The context and conditions in practice are important factors to consider, and understanding the limitations may make expectations more realistic. The presence of ambiguity<sup>145</sup> and lack of self-efficacy<sup>146</sup> may also be significant factors.

Continuous human observation of compliance would not be valid and is unlikely to be affordable<sup>12</sup> or ethical. Other automated options are available but may have limitations including cost-effectiveness<sup>147</sup>. Instead the regular monitoring of infection control practice with clearly defined and agreed compliance outcomes may be feasible and cost-effective. Particularly if it also increases validity by reducing errors, ambiguity and potential variability of scoring.

**Funding** This research was funded by UCLH NHS Foundation Trust.



## References